

VU Research Portal

Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie

Harte, J.M.; Breukink, M.D.

published in

Tijdschrift voor Criminologie
2010

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Harte, J. M., & Breukink, M. D. (2010). Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie. *Tijdschrift voor Criminologie*, 52(1), 52-72.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Objectiviteit of schijnzekerheid?

Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie

Joke Harte & Minke Breukink

In toenemende mate wordt voor het inschatten van toekomstig delictgevaar gebruikgemaakt van instrumenten voor risicotaxatie. Dit artikel geeft een overzicht van de instrumenten die in Nederland in gebruik zijn voor het voorspellen van de kans op toekomstig delictgedrag. Om een beeld te krijgen van de kwaliteit van deze instrumenten zijn studies naar de interbeoordelaarsbetrouwbaarheid en de predictieve validiteit verzameld. De opbrengsten en beperkingen van instrumenten voor risicotaxatie worden gewogen en geconcludeerd wordt dat met de ontwikkeling en implementatie van deze instrumenten een aantal belangrijke verbeteringen is bewerkstelligd, met name voor de behandelpraktijk. Het onderzoek laat echter ook zien dat de mogelijkheid om de kans op recidive goed in te schatten, ook met deze instrumenten beperkt blijft. Het is zaak dat de gedragsdeskundige de rechter hierover goed informeert.

In de afwegingen die een rechter maakt bij het opleggen van een straf of een maatregel, speelt een eventueel gevaar voor recidive een belangrijke rol. Wanneer dit recidivegevaar aanwezig is, en (mede) voortkomt uit een psychiatrische stoornis, kan de rechter besluiten om de verdachte een behandelmaatregel op te leggen om de maatschappij te beveiligen. Voorbeelden van dergelijke maatregelen zijn de terbeschikkingstelling (tbs), de maatregel tot plaatsing in een inrichting voor stelselmatige daders (de ISD-maatregel) en de plaatsing in een psychiatrisch ziekenhuis (art. 37 WvS). Een dergelijke gedwongen behandeling is vaak langdurig, vindt plaats in een beveiligde inrichting en is voor de betrokkene uitermate ingrijpend. Bovendien zijn de kosten van dergelijke behandelingen hoog. De rechter moet er bij oplegging dan ook van overtuigd zijn dat de maatregel echt noodzakelijk is, maar mag tegelijkertijd niet te terughoudend handelen. Want wanneer een delinquent recidiveert met een ernstig delict heeft dit niet alleen directe gevolgen voor de slachtoffers, maar leidt dit ook tot grote maatschappelijke en politieke onrust. Kortom, de strafrechter heeft een nauwkeurige inschatting nodig van het recidivegevaar.

Over de persoon van de verdachte en de kans op recidive laat de rechter zich voorlichten door gedragsdeskundigen. De afgelopen jaren zijn de taxaties van het risico op toekomstig delictgevaar in de gedragskundige rapportages veranderd. In sterk toenemende mate wordt bij deze inschattingen gebruikgemaakt van zogeheten instrumenten voor risicotaxatie. Zowel in de rapportage als ter zitting grijpt de gedragskundige in zijn oordeel vaak terug op de resultaten van deze instrumenten. Maar wat behelzen deze instrumenten precies? En wat is bekend over de

kwaliteit van deze instrumenten? En mag de rechter ervan uitgaan dat het gebruik van deze instrumenten resulteert in juiste inschattingen? In dit artikel worden verschillende methoden besproken voor het inschatten van het risico op toekomstig delinquent gedrag op individueel niveau, waarbij de onderzochte, zijn achtergrond en het eventuele strafdossier bekend zijn. Dit betekent dat methoden voor de analyse van dreigingen of het opstellen van een daderprofiel hier buiten beschouwing blijven. Gekeken wordt welke instrumenten in Nederland in gebruik zijn en wat er bekend is over de kwaliteit van deze instrumenten. Tot slot worden de opbrengsten en de beperkingen van instrumenten voor risicotaxatie tegen elkaar afgewogen en wordt gekeken wat de implicaties hiervan zijn voor de advisering aan de rechter.

Methoden voor risicotaxatie

Lange tijd was het de gewoonte om het recidiverisico van een patiënt puur klinisch in te schatten (Lammers, 2007). Een behandelverantwoordelijke taxeerde het mogelijke gevaar op basis van zijn kennis en ervaring in combinatie met het beeld van de patiënt. Diverse empirische studies laten zien dat deze voorspellingen vaak niet of nauwelijks beter zijn dan toeval (zie onder meer De Ruiter, 2000; Philipse, 2005; De Vogel, 2005). In reactie hierop heeft het wetenschappelijke onderzoek in de forensische psychiatrie zich de afgelopen jaren vooral gericht op het ontwikkelen van instrumenten voor risicotaxatie.

De factoren (oftewel de items) binnen deze instrumenten vormen vaak verschillende schalen of domeinen. Ze zijn geselecteerd op grond van empirisch onderzoek naar de samenhang met recidive en ervaring uit de behandelpraktijk. Een factor kan statisch of dynamisch zijn. Statische factoren zijn kenmerken die niet meer veranderbaar zijn. Een voorbeeld hiervan is de factor *Leeftijd eerste geweld-dadige incident*. Bij dynamische factoren gaat het om kenmerken die veranderbaar zijn en daardoor niet vaststaan, zoals de factor *Ziekte-inzicht*. In de behandelpraktijk zijn vooral de dynamische factoren van belang, omdat ze handvatten bieden voor behandeling. Bij herhaalde meting kunnen scores op dynamische factoren verandering in de problematiek laten zien. Zowel statische als dynamische factoren betreffen meestal kenmerken die bij aanwezigheid de kans op recidive vergroten. Relatief recent is de aandacht voor zogeheten protectieve factoren. Een score op een protectieve factor impliceert bescherming tegen recidivegevaar.

Binnen de instrumenten kan een onderscheid worden gemaakt tussen actuariële risicotaxatie-instrumenten en instrumenten voor een gestructureerd klinisch oordeel. Bij actuariële instrumenten wordt, nadat de items zijn gescoord, een inschatting van het recidivegevaar verkregen door de itemscores op te tellen. Omdat deze instrumenten een sterk generaliserend karakter hebben en geen ruimte bieden voor kenmerken van de specifieke patiënt, zijn deze instrumenten minder geschikt voor de behandelpraktijk. Daar bestaat dan ook een voorkeur voor risicotaxatie-instrumenten, die een zogeheten klinisch gestructureerd oordeel van het recidivegevaar geven. Nadat de beoordelaar alle items heeft gescoord, en zich bewust is van alle relevante informatie, maakt hij zelf een afweging en

komt hij tot een inschatting van het toekomstige delictgevaar. Het betreft dus een klinisch oordeel dat 'gestructureerd' is doordat de beoordelaar zich eerst verdiept heeft in alle relevante informatie met betrekking tot het recidivegevaar van de betreffende patiënt.

Instrumenten voor risicotaxatie in Nederland

In Nederland heeft de ontwikkeling van instrumenten voor risicotaxatie een grote vlucht genomen. Met name binnen het forensisch psychiatrische veld zijn op diverse plaatsen initiatieven ontstaan om instrumenten te vertalen of zelf te ontwikkelen. Aan de hand van de literatuur en met behulp van enkele relevante internetsites¹ is geïnventariseerd welke instrumenten in Nederland in gebruik zijn. Vervolgens is dit overzicht voorgelegd aan het Onderzoekersoverleg Forensische Gedragswetenschappen, een netwerk van wetenschappelijk onderzoekers in het veld van de forensische psychiatrie en psychologie. Dit heeft nog diverse aanvullingen opgeleverd. Vervolgens zijn alleen die instrumenten geselecteerd die (mede) een inschatting geven van toekomstig delictgevaar. Instrumenten die alleen bepaalde (met recidive samenhangende) problematiek in beeld brengen, zijn buiten beschouwing gelaten. Dit impliceert dat instrumenten als de PCL-R (Psychopathy Checklist-Revised; Vertommen e.a., 2002) niet in het overzicht zijn opgenomen.

De 26 instrumenten die op deze wijze verzameld zijn, worden beschreven in tabel 1. De meeste instrumenten betreffen vertalingen van instrumenten uit het buitenland. Relatief veel instrumenten zijn ontwikkeld of vertaald ten behoeve van de Nederlandse tbs-sector en richten zich op het voorspellen van recidive door delinquenten met ernstige psychiatrische problematiek. Dit is verklaarbaar vanuit het feit dat de beslissing rond het opleggen en beëindigen van de maatregel tbs en het verlenen van verlof aan tbs-gestelden gecompliceerd is en gevoelig ligt. Als gevolg van enkele incidenten met tbs-gestelden die recidiveerden tijdens verlof, kwamen verantwoordelijken zwaar onder vuur te liggen en zijn deze beslissingen zowel politiek als maatschappelijk onder grote druk komen te staan (zie Harte, 2007). De behoefte aan ondersteuning bij en verbetering van deze beslissingen is dan ook groot.

Met name vanuit het ministerie van Justitie is op de tbs-sector druk uitgeoefend om tot landelijke consensus te komen wat betreft het instrumentarium. Dit heeft een stevige discussie opgeroepen. Zo is er veel steun voor de HCR-20 (Historical Clinical Risk-20; Philipse e.a., 1999), een instrument dat ontwikkeld is in Noord-Amerika en waarmee internationaal veel onderzoek is verricht. De HKT-30 (Historisch Klinisch Toekomst-30, Werkgroep Risicotaxatie Forensische Psychiatrie, 2003) daarentegen is speciaal voor de Nederlandse tbs-sector ontwikkeld en is grotendeels gebaseerd op diverse bestaande instrumenten. Qua structuur lijkt de HKT-30 veel op de HCR-20, maar bevat meer dynamische items. De Ruiter (2001)

1 Onder meer www.corinederuiter.eu, www.fotres.ch, www.nederlandsjeugdinstituut.nl en www.wodc.nl.

noemde de HKT-30 'een misplaatst poldermodel'. Volgens recent onderzoek (zie Lammers, 2007) zou de HKT-30 echter iets beter presteren dan de HCR-20. Naast het voorspellen van (algemene) recidive door (voormalige) tbs-gestelden zijn ook voor andere populaties instrumenten ontwikkeld, zoals voor jongeren met gedragsproblemen of delinquent gedrag en voor zedendelinquenten. Ook zijn er instrumenten voor het voorspellen van misbruik en mishandeling van kinderen door de ouders en voor het voorspellen van huiselijk en relationeel geweld.

De omvang van de verschillende instrumenten is zeer uiteenlopend. Zo omvat de FP40 twaalf vragenlijsten met in totaal ruim vijfhonderd items en bestaat de RRASOR uit slechts vier items. In vrijwel alle instrumenten worden de items gescoord op een beoordelingsschaal. Zeer verschillend is de wijze waarop het eindoordeel over het recidivegevaar tot stand komt. Bij een aantal instrumenten moet een somscore berekend worden, waarbij soms wordt vergeleken met een *cut-off* waarde. Veel instrumenten vragen om een gestructureerd klinisch oordeel.

Kwaliteit van de instrumenten

Het onderzoek naar de kwaliteit van een instrument voor risicotaxatie richt zich vooral op twee aspecten: de interbeoordelaarsbetrouwbaarheid en de predictieve validiteit.

Interbeoordelaarsbetrouwbaarheid

Voor een goed instrument geldt dat de beoordeling van een casus niet of nauwelijks afhangt van de persoon die het instrument scoort. Verschillende beoordelaars van eenzelfde casus moeten dus tot overeenkomstige beoordelingen komen. In onderzoek naar deze mate van overeenkomst wordt veel gebruikgemaakt van de Intraclass Correlatie Coëfficiënt (ICC; Shrout & Fleiss, 1979). Bij een waarde van 0 is er geen overeenstemming tussen de beoordelaars; bij een waarde van 1 is er perfecte overeenstemming.

In tabel 1 (op p. 56 e.v.) is te zien dat er voor twaalf instrumenten onderzoek is verricht naar de interbeoordelaarsbetrouwbaarheid. In totaal hebben er 21 studies² in Nederland plaatsgevonden met de Nederlandse versie van het instrument. De idee achter de beperking tot Nederlands onderzoek is dat indices voor de betrouwbaarheid en de validiteit van meetinstrumenten, zoals die naar voren komen in buitenlands onderzoek, niet gegeneraliseerd kunnen worden naar de Nederlandse situatie. Een vertaling kan al snel een heel andere betekenis geven aan een begrip, waardoor de eenduidigheid van een item door vertaling sterk kan veranderen. Dit is de reden dat kwaliteitsaspecten van een instrument na vertaling altijd opnieuw moeten worden vastgesteld.

2 Het is niet uit te sluiten dat een enkele studie niet is gevonden. Er vindt momenteel veel onderzoek plaats en bovendien is de ervaring dat de resultaten van de studies regelmatig gerapporteerd worden in ongepubliceerde onderzoeksrapporten.

Tabel 1: *Overzicht instrumenten voor risicotaxatie in Nederland*

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
Best-Index Behavioral Status Index (Reed e.a., 1999)	Psychiatrische patiënten, onder wie tbs-gestelden	Totaal 150 items, verdeeld over 5 gedragsschalen en 1 risicoschaal	Scores worden gesommeerd	Chakhssi e.a. (2009): ICC varieert tussen .66 en .71	Chakhssi e.a. (2009): matige maar significante correlatie met totaalscore HCR-20
CAPI Child Abuse Potential Inventory (Milner, 1994)	Ouders met een kind onder de 18 jaar	Totaal 160 items, omvat Kinder-misbruiksschaal, 3 validiteitschalen en de schalen Ego-strength en Eenzaamheid	Scores worden opgeteld. Bij een cut-off score van 215 op de Kinder-misbruiksschaal wordt het risico als hoog beoordeeld	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
CARE-NL Child Abuse Risk Evaluation – Nederland (De Ruiter & De Jong, 2006)	Potentiële pleegders van kindermishandeling	18 items verdeeld over 4 domeinen: Ouderfactoren, Ouder-kindfactoren, Kindfactoren en Gezinsfactoren	Gestructureerd klinisch	Van der Hoorn (2008): ICC varieert tussen .56 en 1.00	Van der Hoorn (2008): retrospectieve studie, AUC-waarde varieert van .56 tot .78
EARL-20B Early Assessment Risk List for Boys (Koege e.a., 2000; een Nederlandse versie is in voorbereiding, zie www.corinede.ruiter.eu)	Jongens tussen 6 en 12 jaar met gedragsproblemen	20 items verdeeld over 3 schalen: Family items, Child items, Responsivity items	Gestructureerd klinisch	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend

Tabel 1: (vervolg)

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
EARL-21G Early Assessment Risk List for Girls (Levene e.a., 2004; een Nederlandse versie is in voorbereiding, zie www.cornede.ruiter.eu)	Meisjes tussen 6 en 12 jaar met gedragsproblemen	20 items verdeeld over 3 schalen: Family items, Child items, Responsivity items	Gestructureerd klinisch	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
FOTRES Forensische Operationaliseres Therapie Risiko Evaluations System (zie www.fotres.ch)	Delinquenten met ernstige psychische stoornissen	162 schalen en circa 700 items	Risico wordt samengevat in drie som-scores	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
FP40 Forensische Profiel Lijsten (Brand, 2006)	Tbs-gestelden en forensische patiënten in Nederland	De FP40 bevat 12 lijsten, waaronder Seksualiteit, Verslaving, Coping en Netwerk	Gestructureerd klinisch: de honderden items worden teruggebracht naar schaa scores. Deze vormen het profiel	De psychometrische eigenschappen van de FP40 zijn uitgebreid onderzocht door Brand (2006): ICC van de socio-therapieschalen varieert van .64 tot .86	Brand (2006): prospectief bepaalde AUC-waarde = .72. Voor specifieke groepen delinquenten wordt de AUC-waarde hoger

Tabel 1: (vervolg)

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
HCR-20 Historical Clinical Risk-20 (Philipse e.a., 1999)	Personen met een geweld-dadige voor-geschiedenis en mogelijk een psychische of persoonlijkheidsstoornis	3 schalen (Historisch, Klinisch en Risico) met in totaal 10 statische en 10 dynamische items	Gestructureerd klinisch: de beoordeelaar integreert zelf alle informatie tot een eindoordeel (laag, matig of hoog)	De Vogel e.a. (2004): ICC varieert van .58 tot .89 De Vogel & De Ruiter (2005): ICC varieert van .49 tot .82 Hildebrand e.a. (2005): ICC varieert van .56 tot .76	Philipse e.a. (2002): retrospectieve studie, significante AUC-waarden variëren van .78 tot .89 De Vogel e.a. (2004): retrospectieve studie, AUC-waarde varieert van .66 tot .82 De Vogel & De Ruiter (2005): retro-en prospectieve studie, significante AUC-waarden variëren van .73 tot .91 Hildebrand e.a. (2005): retrospectieve studie, AUC-waarde varieert van .62 tot .71 De Vogel & De Ruiter (2006): prospectieve studie, AUC-waarde varieert van .77 tot .86
HKT-30 Historisch Klinisch Toekomst-30 (Werkgroep Risicotaxatie Forensische Psychiatrie, 2003)	Tbs-gestelden in Nederland	Omvat 11 historische, 13 klinische en 6 toekomstitems	Klinisch gestructureerd	Werkgroep Risicotaxatie Forensische Psychiatrie (2003): R = .71-.78 Canton e.a. (2004b): ICC varieert van .78 tot .85 Hildebrand e.a. (2005): ICC varieert van .62 tot .87 Schönberger e.a. (2008): ICC varieert van .61 tot .81	Canton e.a. (2004a): retrospectieve studie, AUC-waarde varieert van .59 tot .73 Hildebrand e.a. (2005): retrospectieve studie, AUC-waarde varieert van .67 tot .73 Schönberger e.a. (2008): retrospectieve studie, AUC-waarde varieert van .45 tot .65

Tabel 1: (vervolg)

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
J-SOAP-D Juvenile Sex Offender Assessment Protocol Dutch version (Bullens e.a., 2005)	Seksueel delinquente jongens van 12 tot 18 jaar	4 schalen: Seksuele drangpreoccupatie, Impulsief antisociaal gedrag, Interventie, Stabiliteit thuissituatie	Integratie kansproportie (totaalscore gedeeld door aantal ingevulde items) en gewogen klinisch oordeel	Bullens e.a. (2005): ICC varieert van .78 tot .98	Geen Nederlands onderzoek bekend
LSI-R Level of Service Inventory - Revised (concept-vertaling door Lammers & Philipse uit 2003)	Wordt nergens expliciet vermeld; in ieder geval beide geslachten	54 items verdeeld over 10 componenten, waaronder Crimineel verleden, Alcohol- en drugsproblemen	Optellen totaalscore	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
RiHG Risicotaxatie-instrument Huiselijk Geweld (Römkens & Van Poppe, 2007; Kuppens & Beke, 2008)	(Potentiële) plegers van huiselijk geweld	20 items verdeeld over de domeinen: Mogelijke pleger, Incidentverloop, Gezinsachtergronden	Bij twee of meer sterke signalen per domein wordt hulp-officier van justitie geadviseerd huisverbod op te leggen	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
RISc Recidive Inschattingsschalen (Programma Terugdringen Recidive, 2004)	Justitiabelen in Nederland, zowel intra- als extramuraal	13 subschalen met in totaal 75 items, zowel statisch als dynamisch	Som van de scores leidt tot profiel met de inschatting van het risico	Van der Knaap e.a. (2007): totaalscore T = .68, schalen T = .43 tot .78	Geen Nederlands onderzoek naar de predictieve validiteit bekend. Op dit moment loopt er een onderzoek van het WODC naar de predictieve validiteit van de RISc (www.wodc.nl)

Tabel 1: (vervolg)

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
RRASOR Rapid Risk Assessment for Sex Offender Recidivism (Hanson, 1997; Philipse e.a., 2001)	Volwassen mannelijke seksuele delinquenten	4 items: Eerdere seksuele delicten, Huidige leeftijd, Geslacht slachtoffer, Relatie met slachtoffer	Scores worden opgeteld tot totaalscore	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
SAPROF Structured Assessment of Protective Factors (De Vogel e.a., 2007)	Volwassen mannelijke geweldsdelinquenten met een psychische of persoonlijkheidsstoornis	17 items verdeeld over 3 schalen: Interne items, Motivationale items, Externe items	Klinisch gestructureerd	De Vries Robbé & De Vogel (2009): ICC varieert van .73 tot .88	De Vries Robbé & De Vogel (2009): retrospectieve studie, AUC-waarde varieert van .71 tot .93
SARA Spousal Assault Risk Assessment (Dutton & Kropp, 2000; Nederlandse versie is in voorbereiding; zie www.corinederuitter.eu)	Volwassen mannelijke plegers van relationeel geweld	20 items verdeeld over 2 domeinen: General violence risk factors, Spousal violence risk factors	Gestructureerd klinisch oordeel (laag, matig of hoog)	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
SAVRY Structured Assessment of Violence Risk in Youth (Lodewijks e.a., 2003)	Jeugdige delinquenten van 12 tot 18 jaar	30 items verdeeld over de schalen: Historische, Sociale Contextuele, Individuele, Protectieve factoren	Klinisch gestructureerd	Lodewijks e.a. (2008a): ICC varieert van .61 tot .86 Lodewijks e.a. (2008b): ICC varieert van .62 tot .86	Lodewijks e.a. (2008a): prospectieve studie, AUC-waarde Historische = niet significant, overige AUC-waarden variëren van .66 tot .86 Lodewijks e.a. (2008b): retrospectieve studie, AUC-waarde: Historische = niet significant, overige AUC-waarden variëren van .65 tot .71

Tabel 1: (vervolg)

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
SORAG Sex Offender Risk Appraisal Guide (Quinsey e.a., 1998)	Seksuele delinquenten	14 items	Scores worden opgeteld	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
SORM Structured Outcome Assessment and Community Risk Monitoring (Hilterman & Bouman, 2004)	Forensisch psychiatrische patiënten die onder begeleiding in de samenleving verblijven	30 items verdeeld over 6 domeinen, waaronder Sociaal netwerk, Klinische factoren, Subjectieve waardering	Klinisch gestructureerd	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
START Short-Term Assessment of Risk and Treatability (Webster e.a., 2006; Gzicht, maart 2006; Eltink, 2006)	Patiënten met psychische problematiek	20 items verdeeld over 7 domeinen. Daarnaast kunnen 2 casusspecifieke items worden toegevoegd	Klinisch gestructureerd	Tot 2010 loopt er bij GGZ Drenthe, locatie Assen een onderzoek naar de interbeoordelaarsbetrouwbaarheid en predictieve validiteit van START	Tot 2010 loopt er bij GGZ Drenthe, locatie Assen een onderzoek naar de interbeoordelaarsbetrouwbaarheid en predictieve validiteit van START
Static-99 (Van Beek e.a., 2001)	Volwassen mannelijke seksuele delinquenten	10 statische items	De scores worden bij elkaar opgeteld	De Vogel e.a. (2004): totale score ICC = .80, eindoordeel ICC = .61 Schönberger e.a. (2008): totale score ICC = .83, eindoordeel ICC = .73	De Vogel e.a. (2004): retrospectieve studie, AUC-waarden zijn veelal niet significant, variëren verder tussen .66 en .71 Schönberger e.a. (2008): retrospectieve studie, AUC-waarden zijn veelal niet significant, variëren verder tussen .63 en .67

Tabel 1: (vervolg)

Instrument	Doelgroep	Inhoud	Eindoordeel	Onderzoek naar inter-beoordelaarsbetrouwbaarheid	Onderzoek naar predictieve validiteit
StatRec (Wartna e.a., 2009)	Alle verdachten van misdrijven van 18 jaar en ouder	6 kenmerken	StatRec is een model dat de kans op recidive berekent	Geen Nederlands onderzoek bekend	Wartna e.a. (2009): retrospectief bepaalde AUC-waarden voor verschillende deelpopulaties varieerden tussen .67 en .85
SVR-20 Sexual Violence Risk-20 (Hildebrand e.a., 2001)	Volwassen seksuele delinquenten met psychiatrische problemen	20 items verdeeld over 3 domeinen: Psychosociale aanpassing, Seksuele delicten, Toekomstplannen	Klinisch gestructureerd	De Vogel e.a. (2004): ICC varieert tussen .48 en .78 Koster e.a. (2006): ICC varieert tussen .42 en .72 Schönberger e.a. (2008): ICC varieert tussen .58 en .78	De Vogel e.a. (2004): retrospectieve studie, significante AUC-waarden variëren van .64 tot .83 Koster e.a. (2006): retrospectieve studie, meeste AUC-waarden niet significant, overige variëren tussen .73 en .78 Schönberger e.a. (2008): retrospectieve studie, significante AUC-waarden variëren van .63 tot .66
VD Vragenlijst Delictgevaarlijkheid (Verhagen & Philipse, 1995)	Tbs-gestelden	47 gedragsitems	Inschatting van de kans op recidive op korte respectievelijk lange termijn	Geen Nederlands onderzoek bekend	Onderzoek (Verhagen & Philipse, 1995) waaruit sterke positieve samenhang bleek met contact met justitie
VRAG Violence Risk Appraisal Guide (Grann e.a., 2000)	Mannelijke gewelds- of seksuele delinquenten	12 items	Totaalscore wordt berekend	Geen Nederlands onderzoek bekend	Geen Nederlands onderzoek bekend
-	Volwassen delinquenten in ambulante forensisch psychiatrische behandeling	Items verdeeld over 12 domeinen	Per domein geeft de beoordelaar een algemeen gestructureerd klinisch oordeel	Alleen van de WaagSchaal Jeugd is onderzoek bekend. De ICC-scores variëren van .73 tot .88 (Van Horn e.a., 2008)	Alleen van de WaagSchaal Jeugd is onderzoek bekend. AUC-waarde varieert van .67 tot .77 (Van Horn e.a., 2008)

De resultaten staan samengevat in tabel 1. Opvallend is dat de interbeoordelaars-betrouwbaarheid van een aantal (schalen van) instrumenten matig blijkt te zijn. Het is zaak om voor deze instrumenten nader uit te zoeken welke items of scoringsinstructies verschillend geïnterpreteerd worden. Een relatief lage ICC kan ook het gevolg zijn van dossierinformatie van matige kwaliteit. Onduidelijk is aan welke eisen (dossier)informatie moet voldoen om tot een betrouwbaar oordeel te komen.

Predictieve validiteit

Het uiteindelijke doel van het instrument is het voorspellen van toekomstig delictgedrag. De mate waarin een instrument hiertoe in staat is, noemt men de predictieve validiteit. Deze kan retrospectief of prospectief worden vastgesteld. In retrospectief onderzoek worden ex-patiënten of ex-gedetineerden geselecteerd die al weer enige tijd vrij in de maatschappij bewegen en van wie recidivecijfers bekend zijn. Voor deze personen wordt het instrument gescoord aan de hand van dossiergegevens uit de periode dat de betreffende persoon op het punt stond om terug te keren in de maatschappij. De predictieve validiteit wordt bepaald door de samenhang tussen de scores op het instrument en de recidivegegevens in kaart te brengen.

Retrospectief onderzoek naar de predictieve validiteit kent verschillende beperkingen. Allereerst moet de onderzoeker zich baseren op (oude) dossiergegevens. Deze zijn niet verzameld voor wetenschappelijk onderzoek en de kwaliteit en volledigheid is niet altijd toereikend. Statische variabelen vragen meestal naar feitelijke historische informatie die relatief betrouwbaar en eenvoudig in een dossier terug te vinden is. Dit geldt in mindere mate voor dynamische variabelen; het scoren hiervan laat meer ruimte over voor subjectiviteit. Een voorwaarde voor een gedegen onderzoek is dat de beoordelaar die het instrument retrospectief scoort, geen wetenschap heeft van de eventuele recidive. De onderzoeker moet het dossier als het ware 'blind' scoren (Van den Brink e.a., 2007). Dit betekent dat dossiers van tevoren gescreend moeten worden op verwijzingen naar eventuele latere recidive.

Het is dan ook aan te bevelen om de predictieve validiteit prospectief te onderzoeken. In dit type onderzoek worden kenmerken van patiënten in het heden verzameld. Het onderzoek loopt als het ware mee met de patiënten. Na jaren worden vervolgens ook de recidivecijfers verzameld. Prospectief onderzoek duurt dus lang en is daardoor lastiger te organiseren. Een belangrijk voordeel is dat in geval van gebrekkige dossierinformatie de beoordelaar navraag kan doen bij de patiënt of diens behandelaar. Bovendien kan de beoordelaar niet beïnvloed worden door kennis over eventuele latere recidive; deze kennis is er eenvoudigweg nog niet.

Een beperking van zowel retrospectief als prospectief onderzoek is dat de gegevens betrekking hebben op een selecte groep patiënten, namelijk diegenen die zich (uiteindelijk) weer vrij in de maatschappij begeven. Naarmate een patiënt als meer recidivegevaarlijk wordt beschouwd, is de kans hierop kleiner. Het effect van deze selectie op de resultaten van onderzoek naar de predictieve validiteit is niet bekend.

Uit tabel 1 blijkt dat 12 van de 26 risicotaxatie-instrumenten onderzocht zijn op predictieve validiteit. In totaal gaat het om 22 studies, waarvan 4 prospectief. Het is gebruikelijk om de predictieve validiteit van risicotaxaties uit te drukken in zogenoemde AUC-waarden (Brand, 2005). Een AUC-waarde van .50 betekent dat de voorspellende waarde gelijk is aan toeval en een AUC-waarde van 1.0 betekent een perfecte voorspelling. Een AUC-waarde tussen .50 en .60 wordt beschouwd als onvoldoende, tussen .60 en .70 als matig, tussen .70 en .80 als redelijk, tussen .80 en .90 als goed en boven de .90 als zeer goed. Van tien instrumenten zijn dergelijke AUC-waarden bekend. In vrijwel al dit onderzoek is de AUC-waarde berekend op het niveau van de items, van de totaalscores op de eventuele domeinen, van de totaalscore en van het eventuele klinisch gestructureerde eindoordeel. In tabel 1 is een samenvatting van deze resultaten opgenomen.

Naast de keuze tussen een prospectieve of een retrospectieve onderzoeksopzet staat de onderzoeker die de predictieve validiteit van een instrument wil bepalen, voor meer beslissingen. Ten eerste moet vastgesteld worden wat precies onder recidive wordt verstaan. Is dit elk contact met justitie? Of zijn alleen ernstige geweldsdelicten relevant? En moet onderzoek naar de predictieve validiteit van een instrument voor zedendelinquenten zich beperken tot recidive van zedendelicten? In een aantal studies uit tabel 1 zijn AUC-waarden voor verschillende typen recidive bepaald. Verder moet de onderzoeker beslissen wat de follow-up-tijd is. Uit de studie van De Vries Robbé & De Vogel (2009) blijkt dat de predictieve validiteit afneemt naarmate de follow-up-periode langer is. De verschillen in opzet van de studies verklaren de soms uiteenlopende resultaten per instrument en maken ook dat een vergelijking tussen instrumenten wat betreft de kwaliteit, of zelfs een keuze van 'het beste instrument', niet goed mogelijk is.

Zoals eerder opgemerkt, zijn er als gevolg van grote druk op de beslissingen met betrekking tot de tbs-maatregel relatief veel instrumenten ontwikkeld in de tbs-sector. Deze instrumenten worden volop gebruikt, mede doordat het ministerie van Justitie heeft verplicht dat een aanvraag voor verlof en de jaarlijkse evaluatie daarvan de resultaten van de HCR-20 of de HKT-30 moet omvatten (zie Ministerie van Justitie, 2007). Het gevolg is dat er de afgelopen jaren veel data verzameld zijn. In combinatie met de discussie over het meest geschikte instrument voor de tbs-sector heeft dit ertoe geleid dat er relatief veel onderzoek is verricht naar de kwaliteit van deze instrumenten.

Opbrengsten

In een relatief korte periode zijn in Nederland vele instrumenten voor het voorspellen van recidive ontwikkeld en in gebruik genomen. Maar wat heeft dit opgeleverd?

'De methode van het gestructureerd klinisch oordeel is van grote waarde voor het voorspellen van het risico van toekomstig (seksueel) geweld in de Nederlandse forensische klinische praktijk', aldus de conclusie uit het proefschrift van De Vogel (2005). Uit haar studies, waarin ze gegevens van patiënten uit de Van der Hoevenkliniek (Utrecht) gebruikt, blijkt dat de HCR-20 het risico van toekomstig

geweld significant beter voorspelt dan de ongestructureerde klinische inschatting. In de studie van De Vogel e.a. (2004) liggen de AUC-waarden voor gewelddadige recidive voor zowel de subschalen voor de totaalscore als voor het gestructureerde klinische oordeel rond de .80. Het ongestructureerde klinische oordeel heeft een aanzienlijk lagere AUC-waarde, namelijk .68. Ook andere studies tonen aan dat risicotaxatie-instrumenten een nauwkeuriger voorspelling geven van het recidive-risico dan het ongestructureerde klinische oordeel (zie Lodewijks e.a., 2008b).

Naast het geven van betere voorspellingen is een ander belangrijk voordeel van risicotaxatie-instrumenten dat zij laten zien hoe een beoordeling van het recidive-gevaar is opgebouwd. Zowel voor een behandelteam, voor de rechter die voorgelicht moet worden, als voor de patiënt zelf wordt inzichtelijk wat in dit specifieke geval de risicoverhogende en de beschermende factoren zijn. Bovendien wordt door herhaalde meting zichtbaar of, en zo ja, op welke gebieden de patiënt vooruitgang heeft geboekt. Volgens De Ruiter (2007) bieden risicotaxatie-instrumenten de patiënt bovendien bescherming tegen willekeur, negatieve tegenoverdracht en blinde vlekken bij de diagnosticus.

Verder stelt zij dat er dankzij risicotaxatie-instrumenten meer uniformiteit in de beoordelingen komt: alle patiënten worden langs eenzelfde meetlat gelegd en zijn in mindere mate afhankelijk van de beoordelaar. Bovendien spreken zij meer 'een gemeenschappelijke taal'. Dit alles wordt met name bereikt door toepassing van de zogenoemde consensusmethode. Dit is een werkwijze die door De Vogel is geïntroduceerd en inmiddels door velen is overgenomen (zie Lammers, 2007). In deze methode wordt het scoren van een risicotaxatie-instrument uitgevoerd door twee getrainde deskundigen die alle informatie over een patiënt lezen en vervolgens onafhankelijk van elkaar het instrument scoren. In een daaropvolgend consensusgesprek komt men tot overeenstemming over de scores en het eindoordeel. Deze werkwijze van intersubjectiviteit zorgt ervoor dat de kans op ongewenste cognitieve vertekeningen (zoals *illusoire correlatie* en *confirmation bias*³) wordt geminimaliseerd. Met name behandelaars zijn niet altijd voldoende objectief en lopen het gevaar beïnvloed te worden door persoonlijke sympathieën of antipathieën jegens patiënten. Hiltermann (2001) concludeert dat het gebruik van risicotaxatie-instrumenten in vergelijking met het ongestructureerde klinische oordeel beter is, goedkoper is, ethischer is en meer rechtsgelijkheid geeft.

Een andere belangrijke opbrengst van de ontwikkeling en implementatie van risicotaxatie-instrumenten betreft dataverzameling voor wetenschappelijk onderzoek. Dankzij de risicotaxatie-instrumenten is het mogelijk om op grote schaal op uniforme wijze data te verzamelen over grote groepen forensisch psychiatrische patiënten. De op deze wijze verkregen datasets zijn van belang voor empirisch wetenschappelijk onderzoek, zowel op het gebied van risicotaxatie, diagnostiek als evaluatieonderzoek.

3 *Illusoire correlatie* is de waarneming van een samenhang die in werkelijkheid niet of in mindere mate aanwezig is. *Confirmation bias* is de neiging om veronderstellingen zo te toetsen dat ze worden bevestigd.

Beperkingen

Dat risicotaxatie-instrumenten veel te bieden hebben, zowel voor de advies- en behandelpraktijk, alsook voor wetenschappelijk onderzoek, is duidelijk. Voor een genuanceerde discussie is het echter ook van belang een aantal beperkingen onder ogen te zien.

Een bezwaar dat regelmatig wordt ingebracht tegen risicotaxatie-instrumenten, is het ontbreken van een theoretisch kader. De werkwijze wordt ook wel gekarakteriseerd als die van een *dustbowl* (zie Van Nieuwenhuizen & Philipse, 2002), oftewel een willekeurig bijeengewaaide hoop los zand. Veel van de nu bekende voorspellers zijn gevonden om de simpele reden dat zij als gegeven in grote administratieve (niet specifiek voor onderzoeksdoeleinden aangelegde) databestanden aanwezig waren, waarvan gebruik is gemaakt bij het ontwikkelen van risicotaxatie-instrumenten. Deze instrumenten omvatten daardoor vrijwel uitsluitend het soort kenmerken waarover bijna altijd wel wordt geregistreerd.

Behalve dat deze kenmerken samenhangen met recidivegevaar, is vervolgens niets bekend over de betekenis van de relatie. Vaak wordt ten onrechte verondersteld dat deze causaal is. Het is goed mogelijk dat bijvoorbeeld variabelen met betrekking tot arbeidsverleden, instabiliteit van relaties en gebrek aan ziekte-inzicht samenhangen met delictgevaar. Maar in plaats van een causaal verband kan het ook een spurieuze relatie betreffen. Het is namelijk goed mogelijk dat al deze variabelen een causale relatie hebben met weer een andere variabele, bijvoorbeeld de psychiatrische stoornis. In dit geval zullen interventies om het ziekte-inzicht te vergroten en werkervaring en sociale vaardigheden te verstevigen niet resulteren in een inperking van het delictgevaar. Voor het ontwikkelen van werkzame interventies is een theoretisch verklarend model dan ook onmisbaar. Verder mag niet uit het oog verloren worden dat er bij een eventuele oplegging van een maatregel sprake moet zijn van recidivegevaar *op grond van een psychiatrische stoornis*. Deze vraag wordt niet beantwoord met behulp van een instrument voor risicotaxatie (zie ook Mooij, 2004, 172-175).

Verder is ook een kritische beschouwing van de voorspellende waarde van de risicotaxatie-instrumenten op zijn plaats. Zijn de voorspellingen van delictgevaar echt zo veel beter dan de klinische inschatting? In de studies die dit onderzochten, is het klinische oordeel vaak nogal gebrekkig geoperationaliseerd. In de eerdergenoemde studie van De Vogel e.a. (2004) definieerden zij het klinische oordeel als de wijze waarop de patiënt de kliniek verliet. Maar ook zonder klinische inschatting als referentiekader is het mogelijk de predictieve waarde van risicotaxatie-instrumenten te bestuderen. Uit tabel 1 valt op te maken dat predictieve validiteit van de instrumenten nog te wensen overlaat. In de meeste gevallen moet deze als matig of redelijk gekwalificeerd worden. Dit betekent dat uiteindelijke recidive relatief vaak niet voorspeld wordt.

In het onderzoek naar de predictieve validiteit ligt de nadruk op het vermogen van de instrumenten om recidive te voorspellen. Het omgekeerde is echter ook van belang, namelijk dat een instrument goed voorspelt dat iemand *niet* zal recidiveren. Wanneer mensen ten onrechte als recidivegevaarlijk worden aangewezen, kan dit resulteren in (voortzetting van) opsluiting en behandeling, terwijl dit

voor de bescherming van de maatschappij niet noodzakelijk is. Over de omvang van deze zogeheten fout-positieven is weinig bekend. Dit geldt overigens evenzeer voor de situatie waarin men alleen op het zogeheten subjectieve oordeel afgaat. Enkele retrospectieve studies naar risicotaxatie-instrumenten geven een indruk van de omvang van deze fout. In het onderzoek van Hildebrand e.a. (2005) werden dossiers van ex-tbs-gestelden van wie de maatregel inmiddels enkele jaren geleden was beëindigd, retrospectief gescoord. Het blijkt dat ongeveer 45 procent van diegenen die volgens risicotaxatie-instrumenten als hoog recidive-gevaarlijk beschouwd moest worden, niet opnieuw was veroordeeld voor enig nieuw strafbaar feit.

De samenvattende resultaten in tabel 1 beperken zich tot de voorspellende waarde op het niveau van subschalen en eindoordelen. Onder meer in de studie van Hildebrand e.a. (2005) is ook op itemniveau gekeken naar de voorspellende waarde. Uit dit onderzoek naar de HCR-20 en de HKT-30 blijkt dat met name items met betrekking tot het al dan niet schenden van voorwaarden en eventuele onttrekking aan toezicht in het verleden in staat zijn toekomstig delictgedrag te voorspellen. Veel items hebben geen of nauwelijks voorspellende waarde. Zoals eerder opgemerkt, betreffen de meeste studies in tabel 1 retrospectieve studies, waarin de onderzoekers zich moesten baseren op dossierinformatie van enkele jaren geleden. Dit is niet gelijk te stellen aan de behandelpraktijk waarin men eventueel aanvullend informatie kan verzamelen. Desalniettemin wordt duidelijk dat de verwachtingen ten aanzien van de voorspellende waarde van de huidige instrumenten niet al te hoog gespannen mogen zijn.

Discussie

Met de ontwikkeling en implementatie van instrumenten voor risicotaxatie zijn belangrijke verbeteringen bewerkstelligd. Deze instrumenten hebben er onder meer voor gezorgd dat de forensische behandeling zich nog meer richt op haar feitelijke doel: het terugdringen van de kans op recidive. Eventuele verschillen van mening hierover tussen de betrokkenen, inclusief de patiënt zelf, komen eerder aan het licht. Bovendien wordt de communicatie vergemakkelijkt.

Het empirische wetenschappelijke onderzoek in de forensische psychiatrie is de afgelopen jaren gedomineerd door het onderwerp risicotaxatie. Opvallend is dat naar de inhoud en de effectiviteit van de forensisch psychiatrische behandeling nog relatief weinig onderzoek is verricht. De vraag doet zich nu voor of het wel ethisch is om een individu een psychiatrische behandeling op te leggen waarvan onzeker is of deze behandeling werkzaam is. Het ziet ernaar uit dat de forensische sector de komende jaren veel zal investeren in het onderzoek naar werkzaamheid van interventies. Routine Outcome Monitoring (De Beurs & Barendregt, 2008), waarbij uitkomstgegevens herhaald worden gemeten, lijkt hiervoor de geëigende onderzoeksmethode. Zowel op individueel als op groepsniveau kan in kaart gebracht worden welk effect de behandeling heeft op toe- of afname van risico- en beschermende factoren.

Voor het herhaald en gestandaardiseerd verzamelen van gegevens over kenmerken die in relatie staan tot (toekomstig) delictgevaar, bieden risicotaxatie-instrumenten met relatief veel dynamische items mogelijkheden. Dit onderzoek moet zich niet beperken tot het beschrijven van relaties, maar zich vooral richten op het vinden van verklaringen. Alleen wanneer bekend is *waarom* patiënten met bepaalde kenmerken een grote kans hebben om te recidiveren, kunnen effectieve behandelprogramma's opgezet worden en kan recidive op inhoudelijke gronden voorspeld worden.

Maar heeft de intrede van de instrumenten ook bewerkstelligd dat de kans op toekomstig delictgedrag beter te voorspellen is? De resultaten van dit onderzoek laten zien dat dit niet waarschijnlijk is. Veel instrumenten zijn (nog) niet door middel van Nederlands onderzoek op betrouwbaarheid en predictieve validiteit onderzocht. Het is overigens de vraag in hoeverre dit met verder onderzoek en adequate instrumenten wel bereikt kan worden (zie ook Brand, 2009). Bogaerts en Nagtegaal (2008) sluiten niet uit dat de wetenschappelijke en statistische grens wat betreft de prestaties van instrumenten al bijna bereikt is. Door middel van trainingen en specifieke instructies is het vaak mogelijk om de interbeoordelaarsbetrouwbaarheid verder te verbeteren. Voor de predictieve validiteit ligt dit lastiger. Uiteindelijke recidive is immers ook altijd in zekere mate afhankelijk van (toevallige) situationele factoren.

Op groepsniveau kunnen de instrumenten een goede indicatie geven van de kans op recidive. Maar bij beoordelingen op individueel niveau, zoals in de advisering aan de rechter of in de besluitvorming over het eventueel verlenen van verlof, ligt dit veelal gecompliceerder. Een extreem hoge score op een redelijk betrouwbaar en valide instrument maakt de kans op recidive aannemelijk. In de praktijk zijn de resultaten echter niet altijd zo uitgesproken. Zowel met als zonder instrumenten voor risicotaxatie is het voorspellen van recidive gecompliceerd en in beperkte mate mogelijk. Het gebruik van scoringslijsten en meetinstrumenten, soms met getallen als uitkomst, kan ten onrechte de indruk wekken van objectiviteit en zekerheid. Het is de taak van gedragsdeskundigen om de rechter in hun advisering goed te informeren over de beperkingen en zorgvuldigheid te betrachten in het (schriftelijk) presenteren van de resultaten.

Er ligt ten slotte ook een discussiepunt voor de strafrechtsdeskundigen onderling. Het betreft de principiële vraag of het aanvaardbaar is om ingrijpende interventies op te leggen op grond van toekomstvoorspellingen. Hierover leven bij juristen veelal andere opvattingen dan bij gedragsdeskundigen en empirisch wetenschappelijke onderzoekers (De Ruiter, 2004). Over het algemeen is de jurist geneigd de fout-positieven (degenen die ten onrechte als risico worden aangeduid) te beschermen. Want ook al behoort iemand tot een hoog-risicogroep, deze ene individuele patiënt kan nu juist die uitzondering zijn die de regel bevestigt. Wanneer we ons uitsluitend baseren op toekomstvoorspellingen ontkennen we de menselijke mogelijkheid tot verandering en daarmee de vrije wil. Dit zou het einde zijn van de rechtsstaat, aldus Buruma (2007). Tegelijkertijd meent Buruma ook dat het strafrecht zijn ogen niet mag sluiten voor wetenschappelijke ontwikkelingen zoals instrumenten voor risicotaxatie. Het is nu aan de strafrechtsdeskundigen om te bepalen welk gewicht de risicotaxaties moeten hebben in de beslissing van

de rechter. Dat in deze discussie ook de betrouwbaarheid en validiteit van de instrumenten een rol moeten spelen, spreekt voor zich.

Literatuur

- Beek, D. van, Doncker, D. de & Ruiter, C. de (2001). *Static-99. Hanson & Thornton, Nederlandse geautoriseerde vertaling*. Utrecht: Forum Educatief, Dr. Henri van der Hoevenstichting.
- Beurs, E. de & Barendregt, M. (2008). *Mogelijkheden voor therapie-effectonderzoek in de TBS-sector. Komen tot een evidence base onder zorgprogramma's*. Den Haag: WODC.
- Bogaerts, S. & Nagtegaal, M.H. (2008). Risk assessment, risk management en nieuwe ontwikkelingen binnen het forensische landschap. In: T.I. Oei & L.H.W.M. Kaiser (red.). *Forensische psychiatrie onderweg*. Nijmegen: Wolf Legal Publishers, 349-357.
- Brand, E.F.J.M. (2005). Een maat voor de kwaliteit van instrumenten voor risicotaxatie. In: M.J. Sjerps & J.A. Coster van Voorhout (red.). *Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht*. Deventer: Kluwer, 429-456.
- Brand, E.F.J.M. (2006). *Handleiding Forensische Profiel Lijsten FP40 v2006*. Den Haag/Utrecht: Dienst Justitiële Inrichtingen/Oostvaarderskliniek.
- Brand, E.F.J.M. (2009). De mechanische en de klinische methode van risicotaxatie. In: J.M. Harte, M.F.M. Verhagen & M.J. Zomer (red.). 'Most probably the best professor in forensic psychiatry'. *Liber amicorum prof. dr. Dick Raes*. Nijmegen: Wolf Legal Publishers, 199-212.
- Brink, R. van den, Drieschner, K. & Harte, J.M. (2007). Onttrekkingen en recidives door tbs-ers tijdens verlof. Nuanceren van een onderzoek in opdracht van de Tweede Kamer. *Tijdschrift voor Criminologie*, 49(1), 68-75.
- Bullens, R.A.R., Horn, J.E. van, Eck, A. van & Das, J. (2005). *J-SOAP D. De Nederlandse vertaling en bewerking van de J-SOAP II (Juvenile Sex Offender Protocol II)*. Utrecht: Forum Educatief.
- Buruma, Y. (2007). Risico-inschatting. Het smalle pad tussen sciëntisme en kwakzalverij. *Delikt en Delinkwent*, 37(4), 350-359.
- Canton, W.J., Veer, T.S. van der, Panhuis, P.J.A. van, Verheul, R. & Brink, W. van den (2004a). De voorspellende waarde van risicotaxatie in de pro Justitia rapportage. Onderzoek naar de HKT-30 en de klinische inschatting. *Tijdschrift voor Psychiatrie*, 46(8), 525-536.
- Canton, W.J., Veer, T.S. van der, Panhuis, P.J.A. van, Verheul, R. & Brink, W. van den (2004b). De betrouwbaarheid van risicotaxatie in de pro Justitia rapportage. Een onderzoek met behulp van de HKT-30. *Tijdschrift voor Psychiatrie*, 46(8), 537-542.
- Chakhssi, F., Ruiter, C. de & Bernstein, D. (2009). Reliability and validity of the Dutch version of the Behavioral Status Index. *Assessment OnlineFirst*, 21 August 2009. Ontleend aan <http://asmnt.sagepub.com>.
- Dutton, D.G. & Kropp, P.R. (2000). A review of domestic violence risk instruments. *Trauma, Violence and Abuse*, 1(2), 171-181.
- Eltink, E. (2006). De START. Short-Term Assessment of Risk and Treatability. *EFN Nieuwsbrief*, mei, 10-11.
- Grann, M., Belfrage, H. & Tengström, A. (2000). Actuarial assessment of risk for violence. Predictive validity of the VRAG and the Historical Part of the HCR-20. *Criminal Justice and Behavior*, 27(1), 97-114.
- Gzicht, 'Risicotaxatie is méér dan de deur naar buiten', maart 2006.

- Hanson, R.K. (1997). *The development of a brief actuarial risk scale for sexual offense recidivism*. Public Works and Government Services Canada.
- Harte, J.M. (2007). Gaan we de tbs-maatregel verlengen of beëindigen? Beschouwing naar aanleiding van een parlementair onderzoek. *MGV. Maandblad Geestelijke Volksgezondheid*, 62, 612-626.
- Hildebrand, M., Ruiter, C. de & Beek, D. van (2001). *Handleiding bij de Sexual Violence Risk-20. Richtlijnen ter beoordeling van het risico van seksueel gewelddadig gedrag*. Utrecht: Forum Educatief.
- Hildebrand, M., Hesper, B.L., Spreen, M. & Nijman, H.L.I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie. Een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R*. Utrecht: Expertisecentrum Forensische Psychiatrie.
- Hilterman, E. (2001). Zekere onzekerheid of onzekere zekerheid. *Sancties*, 303-308.
- Hilterman, E.L.B. & Bouman, Y.H.A. (2004). *SORM Structured Outcome Assessment and Community Risk Monitoring*. Venray/Nijmegen: FPI de Rooyse Wissel/Pompestichting.
- Hoorn, S. van der (2008). *Betrouwbaarheid en validiteit van de Child Abuse Risk Evaluation-Nederland (CARE-NL). Retrospectief onderzoek bij het Advies- en Meldpunt Kindermishandeling*. Amsterdam: Universiteit van Amsterdam.
- Horn, J. van, Bos, M. & Wilpert, J. (2007). De waagschaal. Onze visie op risicotaxatie. *De Waag 'news flash'*, april.
- Horn, J. van, Eisenberg, M. & Wilpert, J. (2008). WaagSchaal Jeugd. Oude wijn in verbeterde zakken? *De Waag 'news flash'*, april.
- Horn, J. van, Wilpert, J., Eisenberg, M. & Mulder, J. (2009). *Handleiding Waagschaal volwassenen, versie 2009*. Utrecht: De Waag.
- Knaap, L.M. van der, Leenarts, L.E.W. & Nijssen, L.T.J. (2007). *Psychometrische kwaliteiten van de Recidive Inschattingsschalen (RISC). Interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit*. Den Haag: WODC.
- Koegl, C.J., Webster, C.D., Michel, M. & Augimeri, L.K. (2000). Coding raw data. Toward understanding raw life. *Child and Youth Care Forum*, 29(4), 229-246.
- Koster, K., Lankveld, J. van & Spreen, M. (2006). Voorspelling van recidive bij zedendelinquenten met behulp van retrospectief gebruik van de PCL-R en SVR-20. *Tijdschrift voor Seksuologie*, 30(4), 204-214.
- Kuppens, J. & Beke, B. (2008). *RiHG. Risicotaxatie-instrument Huiselijk Geweld, versie 2.1*. Arnhem: Advies- en Onderzoeksgroep Beke.
- Lammers, S. (2007). Blijven stoute jongens stout? Taxatie van het recidiverisico in de tbs. *De Psycholoog*, 42, 479-510.
- Lammers, S. & Philipse, M. (2003). *LSI-r concept-vertaling*.
- Levene, K.S., Walsh, M.M., Augimeri, L.K. & Pepler, D. (2004). Linking identification and treatment of early risk factors for female delinquency. In: M. Moretti, C. Odgers & M. Jackson (eds). *Girls and aggression. Contributing factors and intervention principles: Perspectives in law and psychology* (Vol. 19). New York: Kluwer Academic/Plenum, 131-146.
- Lodewijks, H.P.B., Doreleijers, Th.A.H., Ruiter, C. de & Wit-Grouls, H. de (2003). *SAVRY. Handleiding voor de gestructureerde risicotaxatie van gewelddadig gedrag bij jongeren*. Zutphen: Rentray.
- Lodewijks, H.P.B., Doreleijers, Th.A.H., Ruiter, C. de & Borum, R. (2008a). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, 31(3), 263-271.

- Lodewijks, H.P.B., Doreleijers, Th.A.H. & Ruiter, C. de (2008b). SAVRY risk assessment in violent Dutch adolescents. Relation to sentencing and recidivism. *Criminal Justice and Behavior*, 35(6), 696-709.
- Milner, J.S. (1994). Assessing physical child abuse risk. The Child Abuse Potential Inventory. *Clinical Psychology Review*, 14(6), 547-583.
- Ministerie van Justitie. Dienst Justitiële Inrichtingen (2007). *Circulaire. Toetsingskader verlof ter beschikking gestelden*. Den Haag: Ministerie van Justitie.
- Mooij, A.W.M. (2004). *Toerekeningsvatbaarheid. Over handelingsvrijheid*. Amsterdam: Boom.
- Nieuwenhuizen, C. van & Philipse, M. (2002). Risicotaxatie bij zedendelinquenten. Een globaal literatuuroverzicht. *Tijdschrift voor Seksuologie*, 26(1), 70-78.
- Philipse, M. (2005). *Predicting criminal recidivism. Empirical studies and clinical practice in forensic psychiatry*. Nijmegen: Radboud Universiteit.
- Philipse, M., Erven, T. van & Peters, J. (2002). Risicotaxatie in de TBS. Van geloof naar empirie. *Justitiële verkenningen*, 28(8), 77-93.
- Philipse, M., Hilterman, E. & Doren, D. (2001). Tussen mogelijkheid en illusie. Een review van elf risicotaxatie-instrumenten voor seksuele delinquenten. *Tijdschrift voor Criminologie*, 43(1), 2-24.
- Philipse, M., Ruiter, C. de, Hildebrand, M. & Bouman, Y. (1999). *HCR-20. Beoordelen van het risico van gewelddadig gedrag, versie 2*. Nijmegen/Utrecht: Pompestichting/Forum Educatief.
- Programma Terugdringen Recidive (2004). *RISc. Een diagnostisch instrument voor reclasering en gevangeniswezen*. www.justitie.nl/.../Factsheet%20Terugdringen%20Recidive_tcm34-81011.pdf
- Quinsey, V.L., Harris, G.T., Rice, M.E. & Cormier, C.A. (1998). *Violent offenders. Appraising and managing risk*. Washington, DC: American Psychological Association.
- Reed, V., Robinson, D., Woods, Ph. & Erven, T. van (1999). *Behavioural Status Index – Nederlandse versie BSI-D. Een brede risico assessment voor werkers in de geestelijke gezondheidszorg*. Eindhoven: GGzE.
- Römkens, R. & Poppel, J. van (2007). *Bruikbaarheid van het risicotaxatie-instrument huiselijk geweld. Een eerste evaluatie*. Tilburg: IVA.
- Ruiter, C. de (2000). *Voor verbetering vatbaar* (oratie). Amsterdam: Vossiuspers AUP.
- Ruiter, C. de (2001). Een misplaatst poldermodel in de forensische psychiatrie: een reactie op Philipse en Verhagen (2000), *De Psycholoog*, 63, 15-18.
- Ruiter, C. de (2004). Forensisch gedragsonderzoek in strafzaken. *Forensische Expertise*, 30, 50-61.
- Ruiter, C. de (2007). Wat risicotaxatie wel en niet is. Een reactie op A. van den Berg, 'Risico's van risicotaxatie', *Tijdschrift voor Psychotherapie*, 33, 142-149.
- Ruiter, C. de & Jong, E.M. de (2006). *CARE-NL. Richtlijn voor gestructureerde beoordeling van het risico van kindermishandeling*. Utrecht: C. de Ruiter.
- Schönberger, H.J.M., Hildebrand, M., Spreen, M. & Bloem, O. (2008). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie bij seksuele delinquenten*. Utrecht: Expertisecentrum Forensische Psychiatrie.
- Shrout, P.E. & Fleiss, J.L. (1979). Intraclass correlations. Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Verhagen, T. & Philipse, M. (1995). Het voorspellen van het risico van delicttherhaling. *Tijdschrift voor Psychiatrie*, 37(7), 537-552.
- Vertommen, H., Verheul, R., Ruiter, C. de & Hildebrand, M. (2002). *Handleiding Hare Psychopathy Checklist-Revised*. Lisse: Swets en Zeitlinger.

- Vogel, V. de (2005). *Structured risk assessment of (sexual) violence in forensic clinical practice. The HCR-20 and SVR-20 in Dutch forensic psychiatric patients*. Diss. Universiteit van Amsterdam.
- Vogel, V. de & Ruiter, C. de (2005). The HCR-20 in personality disordered female offenders. A comparison with a matched sample of males. *Clinical Psychology and Psychotherapy*, 12(3), 226-240.
- Vogel, V. de & Ruiter, C. de (2006). Structured professional judgment of violence risk in forensic clinical practice. A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime and Law*, 12(3), 321-336.
- Vogel, V. de, Ruiter, C. de, Beek, D. van & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior*, 28(3), 235-252.
- Vogel, V. de, Ruiter, C. de, Bouman, Y. & Vries Robbé, M. de (2007). *SAPROF richtlijnen voor het beoordelen van beschermende factoren voor gewelddadig gedrag*. Utrecht: Forum Educatief.
- Vogel, V. de, Ruiter, C. de, Hildebrand, M., Bos, B. & Ven, P. van de (2004). Type of discharge and risk of recidivism measured by the HCR-20. A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health*, 3(2), 149-165.
- Vries Robbé, M. de & Vogel, V. de (2009). *Assessing protective factors for (sexual) violence. Research results with the SAPROF*. Presentatie op de negende Conference of the International Association of Forensic Mental Health Services, Edinburgh, Schotland.
- Wartna, B., Tollenaar, N. & Bogaerts, S. (2009). StatRec. Inschatting van het recidivegevaar van verdachten van een misdrijf. *Tijdschrift voor Criminologie*, 51(3), 277-295.
- Webster, C.D., Nicholls, T.L., Martin, M.-L., Desmarais, S.L. & Brink, J. (2006). Short-Term Assessment of Risk and Treatability (START). The case for a new structured professional judgment scheme. *Behavioral Sciences and the Law*, 24(6), 747-766.
- Werkgroep Risicotaxatie Forensische Psychiatrie (2003). *Handleiding HKT-30, versie 2002. Risicotaxatie in de forensische psychiatrie*. Den Haag: Dienst Justitiële Inrichtingen.